UNIVERSITY OF EDINBURGH

FACULTY OF SCIENCE AND ENGINEERING

COMPUTER SCIENCE FOUR

PARALLEL ARCHITECTURES

Monday 24 April 2000

2:00 to 3:30

Examiners:

Prof. K. Turner (External)

Prof. G. Brebner (Chairman)

INSTRUCTIONS TO CANDIDATES

Answer any TWO questions.

All questions carry equal weight.

1. (a) Explain how messages might be routed between source nodes and destination nodes in a hypercube network. [*7 marks*]

   (b) Explain how messages might be routed between source nodes and destination nodes in a 3D mesh network such as that used in the Cray T3D. [*7 marks*]

   (c) The designers of a multicomputer system have decided to use either a hypecube or a 3D mesh network to connect together N processors, where N can be varied to suit customers' needs and upgrades. Discuss the pros and cons of these two networks in terms of their scalability with respect to

      i. network diameter
      ii. node degree
      iii. network throughput ($\propto \frac{number\_of\_connections}{average\_message\_distance}$)

   Give values for (i) and (ii) for a 1024-node hypecube and a 1000-node 3D mesh. [*11 marks*]

2. (a) Explain what is meant by the *cache consistency problem* and why it cannot be solved simply by using a write through protocol. [*3 marks*]

(b) Explain how cache consistency is maintained in systems such as the Sequent Balance. [*4 marks*]

(c) Explain the difference between broadcast and directory based protocols and, with the aid of diagrams, show how a simple central directory invalidate protocol operates. [*8 marks*]

(d) Figure 1 shows the state of the directories for a particular cache line in a distributed directory system which uses the SCI (invalidate) protocol. Copy figure 2 into your answer book and annotate it to show what happens when this line is written to in cache 3. [*10 marks*]

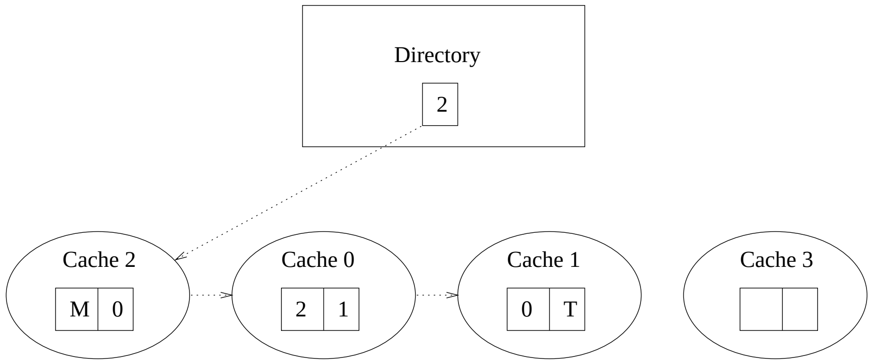

Figure 1: SCI System

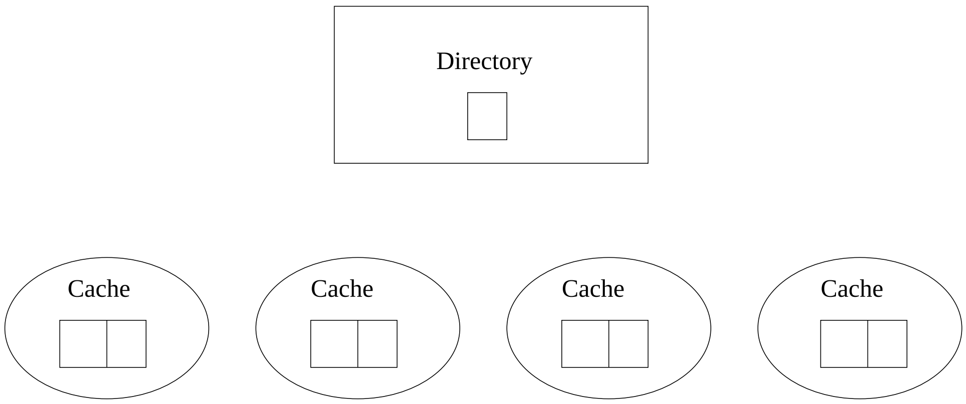

Figure 2: SCI System

3. (a) Describe briefly, with the aid of diagrams, the structure and characteristics of each of a typical UMA, NUMA and COMA multiprocessor architecture. [*7 marks*]

   (b) *Degree of parallelism*, *Speedup*, *Efficiency* and *Utilisation* are all performance metrics used in the context of parallel systems. Explain the meaning of each of these terms. How are they related? [*10 marks*]

   (c) Derive Amdahl's Law as it relates to the speedup offered by an N-processor system.

   Explain Gustafson's objection to the use of Amdahl's Law as a measure of the effectiveness of parallel systems and derive his alternative speedup formula. [*8 marks*]